AD-A115 603 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH F/G 12/1
SCENE ANALYSIS.(U)
DEC 81 W I LUNDGREN
UNCLASSIFIED AFIT/GE/EE/81D-38 NL

END
DATE FILMED
07-82
DTIC

AFIT/GE/EE/81D-38

SCENE ANALYSIS

THESIS

AFIT/GE/EE/81D-38 WILLIAM I. LUNDGREN
1st Lt USAF

AFIT/GE/EE/81D-2

SCENE ANALYSIS

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

by

William I. Lundgren
1st Lt USAF

Graduate Electrical Engineering

December 1981

PREFACE

The thesis effort contained three simultaneous efforts. This paper only presents a theoretical formulation of the scene analysis problem. The formulation is structured so that a mathematically optimal solution can be synthesized. The remaining two topics, analysis of Moshe Horev's thesis effort (Ref 1) and the development of a rectangular to polar coordinate transformation, are reported in AFIT TR-81-1 and AFIT TR-81-2 respectively.

I would like to thank my thesis committee; Dr. Kabrisky, Dr. Lee, and Dr. Maybeck for their open discussions with me. I would also like to thank Dr. J. Hines and Dr. R. Phelps of the Analysis and Signal Processing Group, Radar Branch of the Avionics Laboratory, for use of their computer facilities and for their help in understanding the operating system.

William I. Lundgren

Contents

ABSTRACT

This thesis analyzes the use of a discrete Fourier transform, a template and a frequency filter as scene analysis tools. That analysis leads to a problem formulation that permits mathematical optimization of the frequency filter. The problem is then recast to include the optimization of the transform and template as well. To permit straight forward synthesis, an alternate quadratic cost function is developed, and the minimization of the cost is reduced to a set of linear equations.

The relative merit of norm and dot product correlations and the expected performance of first and second order systems is discussed. Application to reading text is referenced throughout.

SCENE ANALYSIS

I. Introduction

In scene analysis we have a pattern of light intensities that is the result of light reflecting from being emitted from some object or set of objects. The set of objects is called a scene. The task of scene analysis is to locate and/or identify some objects from the pattern given. The task can be extended to three dimensions and include location, identification, rotation and size. That is not considered in this study. The term image, in this study, refers to a specific realization of the light intensity from a scene. That realization is the light intensity at each point of a N by N grid points, and can be considered a vector in R^{N^2} space. The object of interest is referred to as the object and is the digitization of the light emitted or reflected from the object of interest in the scene. The light intensity due to all other objects is referred to as clutter. One popular technique for scene analysis is to take some expected pattern of light intensities for the object and compare that with the light intensities in the image. The expected pattern of light intensities is referred to as the template. In this study the template is taken to be the digitization of the light from the object with no clutter. There is also noise associated with the object that is a result of 3-dimensional rotation of the physical object,

change in lighting or any other possible variations in the scene. We also refer to an average template. That template, as suggested, is the average of a collection of templates of a particular object.

A proposed technique for scene analysis makes use of a discrete Fourier transform (DFT), a template and a frequency filter (Ref:1-3). The author analyzed a particular algorithm of that type and reports that analysis in AFIT TR-81-1. The algorithm analyzed is an algorithm proposed by Moshe Horev, a Major in the Isreali Air Force (Ref:3). The empirical part of that analysis required an algorithm to transform the rectangular coordinate array of light intensities to a polar array of light intensities. A point by point conversion previously used was slow and the author developed an new algorithm to increase the efficiency of that coordinate transformation. That algorithm makes use of the properties of an array processor and a large data storage disk. The details of that algorithm are contained in an AFIT TR-81-2.

The proposed procedure for analyzing a scene is to form an average template, take the DFT of both the unknown template and image, filter both the template and the image in the frequency plane and then compare using either, the norm of the difference between, or the dot product of the two vectors. This study initially starts out formulating a problem statement so that a mathematically optimum frequency filter can be synthesized. That analysis suggests the

plausibility of optimizing not only the filter but the template and the transform as well (the DFT may not be optimum).

A subtle but significant change in the point of view occurs and is detailed in section 2.5. The study starts out investigating the concept of analyzing a scene by using the DFT, a set of templates and a frequency filter. The concept changes until in the final problem statement and solution the concept is to take a set of images (vectors in a vector space, referred to as a data space) whose analysis is known (a human does that), and then to translate the results of that analysis into vectors in a second vector space (one dimension for each result of interest, referred to as the destination space). The final step is to optimize a mapping of the first space into the second. The view is radically different from the starting point but produces an optimization problem that can be solved.

The change is more fundamental than it may at first appear. Normally a problem is approached in terms of a model, in other words, a model is developed based on how the data has been observed to interact (or based on physical laws). The system is then optimized based on that model and the system is run determine if satisfactory performance has been acheived. The approach developed here derives a mapping by requiring that the results of the mapping optimally match the desured (and known) outputs. This optimum mapping is calculated for a limited sample (hopefully representative of

any data that might be encountered) with the assumption that other data will be mapped as desired. That should be the case if the data used to construct the mapping represent a sufficient cross-section of the possible input data. Note that this concept is not limited to the scene analysis problem.

Finally the objective of this optimization and problem formulation is to solve some scene analysis problem rigorously. The problems that were encountered when letter identification was attempted (AFIT TR-81-1) and the resulting frustration prompted this study so that a working solution to some scene analysis problem could be synthesized. The attempt is to start out with a workable problem and to develop insights into the characteristics of this problem formulation and then theorize about how to approach the more complicated problems using those characteristics. The simple problem that is discussed is the reading of uppercase handprinted text. That is even further simplified to the problem of identifying isolated handprinted uppercase letters. The key is that scene analysis is a very complicated subject and the effort here is formulate a simple solvable problem.

II. THEORY

The first idea developed is to optimize the frequency filter. A formal statement of that objective is given in section 2.1. The assumptions to make the mathematical statement simple and clear are discussed in section 2.2. Sections 2.3.1 and 2.4.1 state the optimization problem and simplify the equations as much as possible. That statement suggest it might not be much more difficult to optimize not only the filter but the template and the transformation as well (the DFT may not be optimum). That problem is developed in sections 2.3.2 and 2.4.2. The results of optimizing all three elements of the correlation procedure suggest elimination of the correlation, template, and filter completely and optimizing only a mapping. The remainder of the chapter approaches the problem from that new point of veiw.

2.1 Problem Statement

The problem considered in this chapter is the task of locating and indentifying any one of T objects in a cluttered scene. The particular problem referenced throughout this chapter is reading text. A restriction on the problem is that the text is made up of only uppercase block letters. That problem has 26 classes of objects. A set of test objects is also referenced in the text. That set is made up

of C samples from each class of objects. The statement of the problem is

(1) There are T classes of objects that are of interest.

(2) Given are C samples from each class of objects.

(3) Given are P pages of text with known letter location identification.

(4) One problem is to learn to identify isolated letters.

(5) A second task is to learn to locate the center of letters in text.

(6) A third problem is to identify letters that have been located in text.

(7) A practical but questionable assumption is made; that is, it is assumed that the text given is representative of the of the real text that will be encountered and results from the text given will extrapolate to other text. In other words, the algorithm is trained on a limited set and expected to perform on the universal set.

The problem statement seperates the location and identification task. That seperation may not be necessary but it is an integral part of this problem statement. The assumption is made because putting both into a single cost function will compromise each. The joint cost function is much more complex to formulate and we are looking for a simple, solvable, optimizable problem statement.

2.2 Points of Discussion

2.2.1 Generalized Correlation: Norm vs Dot Product

The norm and the dot product do not give the same result when used as a measure of correlation. A simple example can demonstrate that idea. The templates for the two classes

might be

$$T(1) = (5,2,1)$$

$$T(2) = (3,1,2) \qquad (1)$$

and the unknown is

$$UN = (4,1,1.5) \qquad (2)$$

Notice that the vectors are 3 dimensional and that they can be considered intensities at 3 pixels. The templates could represent a light gradient that fades off to the right and one that fades then brightens again. These patterns would be easily recognizable by a human. The unknown would be an element of the second class. The results of the norm and the dot products between the unknown and the templates are

$$||UN - T(1)|| = 2.25$$

$$||UN - T(2)|| = 1.25 \quad *$$

$$\langle UN, T(1) \rangle = 23.5 \quad *$$

$$\langle UN, T(2) \rangle = 17 \qquad (3)$$

The norm decision is to pick T(2) but the dot product decision is to pick T(1). Notice that the minimum norm (difference) and the maximum dot product (alignment) are measures of likeness. The first thought is to normalize the vectors since that will make the dot product exactly a measure of alignment. The results are then

$$||NT(1) - NUN|| = 0.2102 \quad *$$

$$||NT(2) - NUN|| = 0.2252$$

$$\langle NT(1), NUN \rangle = 0.9779 \quad *$$

$$\langle NT(2), NUN \rangle = 0.9746 \qquad (4)$$

where the N indicates the vectors have been normalized. The decisions are now consistent with each other but not what was intended. The norm and dot product correlations here are at best inconclusive.

This example points out some problems with generalized correlations:

(1) Certain components may dominate the generalized correlation.

(2) The norm and dot product correlations may not give the same results.

(3) The norm of the individual vectors can be important in the decision making (if the vectors are not normalized) for norm correlation (relates to 1).

(4) Alignment is dominated by the large values (relates to 1).

Note that a problem is already developing; there is a discrepency between the results of the generalized correlations and the intended results. This is discussed in more detail in section 2.5. The purpose of this discussion is not to select either the norm or the dot product correlations but to motivate the investigation of both. The relative merit of each will be discussed later.

2.2.2 Transform

There has been a lot of effort to find a scene analysis algorithm via the use of the DFT. Certian characteristics of the human visual process suggest that such a transform might be used by humans (ref:3). There is no mathematical proof that the DFT is optimum or even necessary. The discussions in 2.3 and 2.4 consider both the DFT and a general transform.

The DFT is a pure rotation in complex space. The general will also be confined to be a pure rotation. That is done to maintain the same structure as we are presently using. Later in the chapter we will loosen that restriction.

2.2.3 Preprocessing

The two steps identified do not completely define the problem. Preprocessing may include many things, such as clutter supression, edge enhancement, or any number of other things. That is often the desired result of filtering in the frequency domain (ref:2). The particular preprocessing referred to consist of using a threshold algorithm to black out the background, use a first order moment to center the letter (for identification only) and normalizing the letters. The normalization might consist of setting all non-zero (others set to zero by the threshold algorithm) components to one or it might set the magnitude of the N^2 dimensional vector to one. Preprocessing is not considered in this development, but is discussed in section 2.7. It may seem strange that preprocessing is not included but recall that the first effort will be to formulate a solvable problem. Preprocessing can only be indirectly measured by the cost function and adds tremendously to the complexity. For now, we will assume that the preparation has enhanced letter identification as much as possible.

2.2.4 Non-Linear Filters

The algorithms discussed in this chapter are either norm

or dot product correlations of linear functions. There is nothing that says that a linear algorithm will be able to resolve the scene alalysis problem, in fact the solution may require a much higher order solution. We start out considering a linear problem and later generalize to higher order systems.

2.2.5 Notation

Throughout the rest of this chapter notation is used to try and seperate the various vectors, cost functions and generalized correlations used. Two main catagories of generalized correlations are considered. The first is the dot product correlation and is represented by CDP. The second is norm correlation and is represented by CN. The letters p and t are reserved to indicate classes, there are T classes. There are C samples in each class and the letters c and r are sample indicies and are represented as arguments.

The generalized correlations return T values, one for each class of interest. Those values can be considered components of a T dimensional vector. The indicies corresponding to the components of the vector are expressed as superscripts or subscripts. For example CN^t is the t^{th} component of a T dimensional vector, and I^{kl} is the $(k,l)^{th}$ element of a N^2 dimensional image vector.

Indicial summation notation is used. That notation implies summation over indicies when the same letter is used as superscripts and subscripts. For example

$$I^{kl}(p,r)\ F_{kl}(t) = \sum_{k}\ \sum_{l}\ I^{kl}(p,r)\ F_{kl}(t) \qquad (5)$$

Summations over arguments are still explicitly expressed. The letters i, j, k, l, m and n are summed from 1 to N unless otherwise specified.

There are two stages to calculating the cost for the optimization problem. First the unknown vector is compared with the T templates. The second step is to combine the T values returned (referred to as the dot product or norm correlation values) in to a single cost.

2.3 Indentification

In section 2.3.1 we will set up the problem to optimize the frequency filter used when approaching identification via the DFT. That can be expressed mathematically as

$$CN(t) = ||FFUN - FFAT(t)||^2 \qquad (6)$$

where CN is the norm correlation, FFUN is the filtered DFT of the unknown vector and FFAT(t) is the filtered DFT of the template for the t^{th} class of vectors. Note that the generalized correlation can be done in the transform space since the transform is a pure rotation (see the Appendix). The dot product correlation is

$$CDP(t) = \langle FFUN,\ FFAT(t) \rangle \qquad (7)$$

where the brackets indicate a dot product.

The identity of the letter is chosen as the t corresponding to the smallest value of CN(t) or/and the largest value of CDP(t). The norm correlation can be

expressed in terms of one of the test letters, that is

$$CN^t(r,p) = ||FFI(r,p) - FFAT(t)|| \tag{8}$$

where FFI(r,p) is the filtered DFT of the p^{th} element of the r^{th} class of objects and CN(t,r,p) is the norm correlation of that vector with the t^{th} template. The dot product correlation can be written

$$CDP(t,r,p) = \langle FFI(r,p), FFAT(t) \rangle \tag{9}$$

where CDP(t,r,p) is the dot product correlation value of the p^{th} element of the r^{th} class of templates with the t^{th} template.

Now, one cost for the norm correlation could be

$$COSTN = \sum_{t=1}^{T} \sum_{r=1}^{C} CN(t,t,r) + \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{r=1}^{C} (1 - \delta_{rt}) \frac{A}{CN(t,r,p)} \tag{10}$$

where δ_{rt} is the Kronecker delta function and COSTN is the cost of the norm correlation. Note that the terms in the first sum are just the norm correlation of the elements of each class with the template for the class to which they belong. Those are the norm correlation values we want small. The terms in the second sum are inverse of all other norm correlation values, correlation values that we want large. That means that we want the second sum small as well. Therefore minimizing this cost functional will give us an

optimal solution in which we might be interested.

A cost function for the dot product correlation could be

$$COSTDP = \sum_{t=1}^{C} \sum_{r=1}^{T} \frac{A}{CDP(t,t,r)}$$

$$+ \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{r=1}^{C} (1 - \delta_{pt}) CDP(t,p,r) \quad (11)$$

We want the dot product correlation values in the first sum large and in the second sum small. Again we want to minimize this cost functional to get an optimal solution.

2.3.1 Filter Optimization

A cost function has been developed. Generalized correlation values are to be used in calculating the cost. Expressions for the norm and dot product correlations are given in equations (8) and (9). This section will expand and rearrange those expressions. What we will see is, it may be reasonable to combine the transform, the filter and template so that all can be simultaneously optimized. First we will expand the norm correlation. That is

$$CN(t,p,r) = \sum_{m=1}^{N} \sum_{n=1}^{N} |FFI^{mn}(r,p) - FFAT^{mn}(t)|^2 \quad (12)$$

where the norm in equation (8) has been expanded into a sum of squares of the components. The filtered DFT of the image and template can be expanded by substituting

$$FFI^{mn}(r,p) = F_{(mn)} \, FC_{(mn)kl} \, I^{kl}(p,r) \quad (13)$$

and (the pararenthesis specifies no summation)

$$FFAT^{mn}(t) = F_{(mn)}\, FC_{(mn)kl} AT^{kl}(t) \qquad (14)$$

where F is the frequency filter to optimize, I(p,r) is the r^{th} element from the p^{th} class of images, FC is the array of DFT coefficients (see the Appendix) and AT(t) is the template for the t^{th} class of images. Then

$$FCN(t,p,r) = \sum_{m=1}^{N} \sum_{n=1}^{N} [F^{mn}\, FC_{mnkl}(I^{kl}(p,r) - AT^{kl}(t))] \times [F^{mn}\, FC_{mnij}(I^{ij}(p,r) - AT^{kl}(t))]^* \qquad (15)$$

where the summation over m and n is over all four terms simultaneously. The complex conjugate can be taken inside the brackets (recall the image, template and filter are real), then the equation can be rewritten

$$CN(t,p,r) = \sum_{m=1}^{N} \sum_{n=1}^{N} (F^{mn})^2\, FC_{mnkl}\, FC_{mnij}^* \times [(I^{kl}(p,r) - AT^{kl}(t))(I^{ij}(p,r) - AT^{ij}(t))] \qquad (16)$$

The term in the square brackets can be expanded and the summations carried over each term individually to give

$$CN(t,p,r) = [(F^{mn})^2\, FC_{mnkl}\, FC_{mnij}^*]\, I^{kl}(p,r) I^{ij}(p,r) - 2[(F^{mn})^2 FC_{mnkl} FC_{mnij}^* AT^{ij}(t)] I^{kl}(p,r) + [(F^{mn})^2\, FC_{mnkl}\, FC_{mnij}^*\, AT^{kl}(t)\, AT^{ij}(t)] \qquad (17)$$

From this formulation it is apparent that the norm correlation is a second order correlation in the data (the data is the unknown image). Assuming for the moment that the

optimum filter is known, the summations inside the brackets of equation (17) are completely determined and can be replaced by single arrays. That is

$$CN(t,p,r) = Kl_{klij}(t)\ I^{kl}(p,r)\ I^{ij}(p,r) + K2_{kl}(t)\ I^{kl}(p,r) + K3(t) \qquad (18)$$

where the coefficients Kl, K2(t) and K3(t) are determined by the summations in the square brackets of equation (17).

The dot product can be expanded by substituting equations (13) and (14) into (9), to give

$$CDP(t,r,p) = \sum_{m=1}^{N} \sum_{n=1}^{N} [F_{mn}\ FC_{mnkl}\ I^{kl}(p,r)] \times [F^{mn}\ FC_{mnij}{}^{*}\ AT^{ij}(p,r)] \qquad (19)$$

The summations can be interchanged to give

$$CDP(t,p,r) = \sum_{m=1}^{N} \sum_{n=1}^{N} [(F^{mn})^2\ FC_{mnkl}{}^{*}\ AT^{ij}(t)] \times I^{kl}(p,r) \qquad (20)$$

Again the term in the brackets is completely determined, assuming the optimum filter has been found, and can be replaced with a single array. We get

$$CDP(t,p,r) = K4_{kl}(t)\ I^{kl}(p,r) \qquad (21)$$

Clearly this is a first order equation in the data.

Two important things are brought out by this analysis. The first is that the norm correlation is second order while

the dot product correlation is first order. That would lead one to expect the norm correlation to have more potential than the dot product correlation. The second thing to note is that it may be as simple to optimize all four arrays K1, K2(t), K3(t) and K4(t) as it is to optimize just the filter.

2.3.2 General Optimization

The last section points out that equations (16) and (19) could be used to optimize the templates, filters and the transform coefficients. Since they are multiplied it is equivalent to optimizing the combined arrays given in equations (18) and (21).

The two cost functions developed are exactly the opposite. Forthegeneral optimization it is convenient to eliminate one. The question we need to resolve is which to use. It turns out that that is not difficult.

Wechoose the dot product type of cost function. The reason is because a difference between 2 vectors must be small for the norm type of correlation function to work. That would mean that a single correlation function that had input 2 vectors where the second is the first with a constant added to each component, cannot possibly return a smallvalue in both cases. The dot product correlation does not suffer from that problem.

The generalized optimization problem is then

$$COST(t,p,r) = \sum_{p=1}^{T} \sum_{r=1}^{C} \frac{\Lambda}{CG(r,r,p)}$$

$$+ \sum_{p=1}^{T} \sum_{r=1}^{C} \sum_{t=1}^{T} (1-\delta_{rt}) CG(t,r,p) \qquad (22)$$

and the generalized correlation is given by

$$C1G(t,r,p) = K1_{kl}(t)\ I^{kl}(p,r) + K2(t) \qquad (23)$$

for a first order transformation and

$$C2G(t,p,r) = K3_{klij}(t)\ I^{kl}(p,r)\ I^{ij}(p,r) + K4P_{kl}(t)\ I^{kl}(p,r) + K5(t) \qquad (24)$$

for the second order system. Notice that this could be carried to any order system.

An interesting note is that optimizing a frequency filter for dot product identification is a restriction of the more general problem of optimizing the templates in the spatial domain. Equation (23) is just a dot product in the spatial domain with a constant added. The constant will have a material effect on the decision and cannot be assumed to be zero. The constant arrays $K1^{kl}(t)$ can then be viewed as optimum templates. We showed in the last section that the dot product correlation generalized to this form.

2.4 Location

All the ideas developed in the last section can be applied to the location problem by simply stating the problem in the same format. We will do that in this introduction and the remainder of the development will be straight forward.

The classes have to be established. To do that we need

only find the number of possible locations in the image and establish T' classes. We then pick out C' samples for each class. We now have the saples we need and will be able to use the same equations as developed in section 2.3.

2.4.1 Optimal Filter

Everything stays the same as last section only the classes have been changed. The equations won't be duplicated here. The key concept to note is that the procedure is exactly the same in either case only the data used to produce the optimal filter will change.

2.4.1 General Optimization

Again the equations that we use are exactly the same as for identification, only the sample set has been changed. We can use various ansformations, first, second, third order and so on.

2.5 Alternate Point of View

The motivation for the seperation of location and identification can be put on firmer ground. We might attempt to form T x T' classes, where T is the number of objects of interest and T' is the number of possible locations. That would give optimum location and identification simultaneously except that the previous dimensions of the output vector were T and T', the dimension of this new problem is TxT'. That makes the whole problem more difficult by just increasing the dimension. The other problem is that we may not be getting

what we want. First if we are getting what we want. First if we are viewing a whole page of text simultaneously that will make the number of locations extremely high and secondly there will be multiple locations for the same letters. That cannot be directly approached by this problem formulation. In any case what we probably want to do is look at a relatively small area, locate a letter and then determine the identity of that letter. That specifically requires a seperate location and identification algorithm.

The results of the last two sections suggest a different way of looking at the scene analysis problem. Previously we transformed the data, filtered the data and then did a generalized correlation with some template. That approach is to theorize about a system. develop that system and then check to see if the results are what is desired.

We found in the last two sections that the above approach reduces to first and second order transformations. Since optimizing the first and second order transformations is equivalent to optimizing the template, filter and transformation. That suggest a new point of veiw. We have established a data space. We have established a data space. We can also establish a destination space. That space contains the results of interest to us. We then optimize an R^{th} order transformation between the two. To do that we make use of samples from the data space whose analysis is known. The analysis is done by a human.

The key thing to realize is that in the original point

of view we were attempting to develop a transformation based on our knowledge of the data space, but in the new point of view we are only concerned that the results of the transformation be of use to us. In other words, the only structure that we put on the transformation is the order, and that is only to make the optimization manageable.

2.6 Probablistic Point of View

This section breifly discusses the problem from a probablistic point of view. There are some valuable insights that can be developed from this point of view.

The first step in laying out the problem from the probablistic point of view is to establish a sample space. The events that are of interest are physical realizations of hand printwd uppercase letters. A sample space $_1$ can be defined that consist of all hand printed uppercase letters. Each of these letters, when digitized, can result in many different N^2-tuples of numbers. The N^2-tuple of numbers is the result of digitizing the light intensity at each intersection of a N by N grid of lines covering the letter. The noise associated with the digitization can be taken to be elements second sample space Ω_2. Those spaces are

$$\Omega_1 = \{\omega: \omega \text{ is a hand printed uppercase letter}\}$$
$$\Omega_2 = \{\omega: \omega \text{ is a sample of the noise of digitization}\} \quad (25)$$

The total sample space can be taken to be the cartesion product of those two spaces

$$\Omega = \Omega_1 \times \Omega_2 \quad (26)$$

The assumption made here is that the subspace Ω_2 can be ignored because its effect is relatively small. That assumption can be translated into more physical terms. Given a handprinted uppercase letter and given the digitization of that letter, the displayed digitized letter can easily be recognized as the handprinted letter itself. This is not in general true for all digitized data but we will assume that the digitizer being used is of high quality.

The probability space triple can be set up as

$$(\Omega_1, \mathcal{S}, p,) \tag{27}$$

where Ω_1 is the sample space discribed, $\mathcal{S}$ is the collection of sets on that space that are measureable and form a sigma algebra, and p is the probability measure on those sets. The sigma algebra can be taken to be all sets made up of a single letter, all unions of those sets, the null set and the entire space. The probility triple has been completely defined and the analysis can proceed.

The next step is to define a random vector that maps an element of Ω_1 into R^{N^2}. That random vector is represented by the sequence $x_i(\cdot)$ where $x_i(\cdot)$ is the random variable associated with the i^{th} cooridinate of the vector in R^N (equivalent to $I^{kl}(p,r)$ where $i=(k-1)N+1$). The sample space Ω_1 is divided into 26 classes, and those classes are defined so that each includes all letters of one type. That is

$$C_1 = (w: w \text{ is the result of handprinting an A})$$
$$C_2 = (w: w \text{ is the result of handprinting an B}) \tag{28}$$

and so on. The element of C_1 are referred to as w(i,•). One final set of vectors are needed. The vectors are the vectors of interest in the destination space R^{26}. We want members of C_i to map into the vector $S^j(i)$ where the argument refers to the class to which the vector corresponds and the argument j refers to the component of the vector. We set

$$S^j(i) = \delta_{ij} \tag{29}$$

where δ_{ij} is the Kronecker delta function. The mapping from R^{N^2} to R^{26} can be of any order but for the time being is taken to be linear. The mapping is represented by an array, $F_i(j)$ where i ranges from 1 to N^2 and j ranges from 1 to 26.

The mapping is then

$$y(j,\bullet) = F_i(j)x^i(j) \tag{30}$$

where y(j,•) is a random vector in the destination space and $s^i(\bullet)$ is a random vector in the data space.

The cost function (similiar to equation (22)) is

$$\text{COST} = E\left\{\frac{A}{y(p,w(p,\bullet))} + \sum_{k=1}^{T} y(k,w(p,\bullet))(1-\delta_{pk})\right\} \tag{31}$$

where the expectation is over all omega. Any number of cost functions might work, this one clearly measures what we want. It maximizes the component of the destination vector that corresponds to the true identity (first term) and minimizes all other components. In the next section we consider a quadratic cost function so that we can synthesize a solution.

The probabilistic statement can be reduced to the

deterministic statement as follows. Take $_i$ to consist of T classes of C sample vectors each. Elements of C_t will be denoted w(t,c) where t indicates the class and c indicates which element of that class. We then make the assumption that the probability of occurance of each w(t,c) is $(CT)^{-1}$, that is each element has an equal probability of occurance. The expectation can then be expressed as a sum, or

$$COST = \left\{ \sum_{c=1}^{C} \sum_{t=1}^{T} \frac{A}{y(t,w(t,c))} + \sum_{k=1}^{T} y(k,w(t,c))(1-\delta_{tk}) \right\} \frac{1}{CT} \tag{32}$$

the summations can be rearranged to get

$$COST = \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} \frac{A}{y(t,w(t,c))} + \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{k=1}^{T} (1-\delta_{kt}) y(k,w(t,c)) \tag{33}$$

and that is the same as equation (22) except it has the additional constant factor $(CT)^{-1}$.

The linear condition on the mapping is very limiting. It means that the decisions are based on the value of individual components of the random vector compared with that same component of other vectors. The problem is that no cross correlations between various components of a single random variable can be used. The following paragraph will attempt to clarify why that severely limits us and why higher order solutions might be of use.

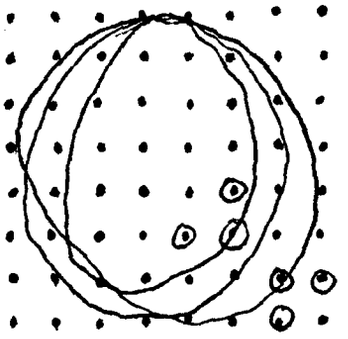
First we will attempt to compare an O and Q using a linear mapping. For simplicity of discussion we will assume the letters have been prepared by a threshold device that has

set all values to either zero or one. We consider a single component of the random variable. The probability that that component will be 1 may be the same for all classes. If that is so there will be absolutely no information of value in that pixel. At the other extreme, there may be a pixel that is always 1 for a given class and always 0 for all other classes. In that case that pixel would contain all the necessary information to identify the single class whose value is 1. How can that be applied to an O and a Q. One would expect that the only reliable difference would be on the lower right side. To identify the Q one might expect that some points toward the center from the circle and some toward the lower right corner from the circle might be used (see figure 1a). It might then be possible that a O that was not quite on center would be bright at some of these points. What we would like is to require that points on a diagonal and some distance apart should be simultaneously large, but that requires a second order mapping (see figure 1b). We would require that a product of pixels connected by the lines in figure 16b be large for the letter to be a Q. The decision would be such that a circle would make the letter an O with a Q, say, 30% lower and then if the diagonal line exsist it will push the Q correlation higher than the O correlation.

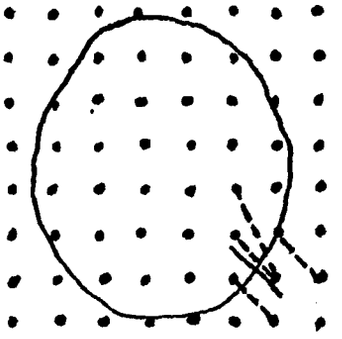
The value of the higher order system is apparent. A third order system can look for curves, straight lines, intersections or continuity. Continuity of curves might be useful in distinguishing between an O and a C for example.

Figure 1

Characteristics of O and Q



a. First Order Pixels



b. Second Order Pixel Pairs

The extension to second order systems is discussed in section 2.3.2. The cost function will remain the same, only the transformation will change.

Two important concepts have been discussed in this section. First, reasonable assumptions takes us directly from the probablistic formulation to the deterministic formulation. The second point is that a linear system severely limits the decision making ability of an algorithm.

2.7 A Solution to Isolated Letter Identification

The large dimensions of the problem and the non-quadratic cost function make the synthesis of a solution very difficult. It may be useful to simplify the problem statement one step further to have a synthesizeable solution. We do that in this section. There is a compromise, we need a quadratic cost function. The compromise will be discussed in the next section.

2.7.1 Quadratic Cost Function

The vector we want to optimize is $y(t,w(p,c))$ where $w(p,c)$ is the c^{th} element of the t^{th} class of vectors. What we really want is all vectors of the form $y(t,w(t,c))$ to be large and all elements of the form $y(t,w(p,c)), t \neq p$ to be small. That is well represented by the cost function given in equation (22). As an alternate we might require that the vectors of the form $y(t,w(t,c))$ to be close to 1. That is a compromise since that is requiring that two vectors that only differ by a constant both map to one, and that uses up cost

on a objective in which we really have no interest. Nonethe-less it does result in a solvable problem. That compromise might not be bad if the images are "normalized". The quadratic cost function is

$$\text{COST} = \frac{1}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} \left\{ y(t,w(p,c)) - \delta_{tp} \right\}^2 (1+B\delta_{tp}) \quad (34)$$

This cost function will drive vectors of the form y(t,w(t,c)) to 1 and vectors of the form y(t,w(p,c)),t≠p to zero. It will put a weighting of 1+B on driving vectors of the form y(t,w(p,c)) toward 1.

2.7.2 Expansion of Cost Function

In order to get the solution to the equation we must first expand the cost in terms of the transform coeffiients, F. That is we substitute

$$y(t,w(p,c)) = F_{kl}(t) I^{kl}(p,c) \quad (35)$$

into the cost function, equation (34). That gives us

$$\text{COST} = (TC)^{-1} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} \left\{ F_{kl}(t) I^{kl}(p,c) - \delta_{tp} \right\}^2 (1+B\delta_{pt}) \quad (36)$$

expanding the square and carrying over the last term gives

$$\text{COST} = \frac{1}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} \left\{ F_{kl}(t) I^{kl}(p,c) \right\} \left\{ F_{kl}(t) I^{kl}(p,c) \right\}$$

$$- \frac{1}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} F_{kl}(t) I^{kl}(p,c) (\delta_{tp} + B\delta_{tp}^2)$$

$$+ \frac{1}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} \left\{ (\delta_{tp})^2 + B(\delta_{tp})^3 \right\}$$

$$+ \frac{1}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} \delta_{pt} \left\{ F_{kl}(t) I^{kl}(p,c) \right\} \left\{ F_{kl}(t) I^{kl}(p,c) \right\}$$

$$+ \frac{1}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} F2(t)^2 \, (1+B\delta_{pt})$$

$$- \frac{2}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} F2(t) \delta_{pt} (1+B)$$

$$+ \frac{2}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} Fl_{kl}(t) I^{kl}(p,c) F2(t)$$

$$+ \frac{2B}{TC} \sum_{t=1}^{T} \sum_{p=1}^{T} \sum_{c=1}^{C} Fl_{kl}(t) I^{kl}(p,c) F2(t) \delta_{pt} \qquad (37)$$

The summations can be interchanged and the delta functions summed over to give

$$COST = \sum_{t=1}^{T} F_{kl}(t) F_{kl}(t) \left\{ \frac{1}{TC} \sum_{p=1}^{T} \sum_{c=1}^{C} I^{kl}(p,c) I^{ij}(p,c) \right\}$$

$$+ \frac{1}{T} \sum_{t=1}^{T} F_{kl}(t) F_{ij}(t) \left\{ \frac{1}{C} \sum_{c=1}^{C} I^{kl}(p,c) I^{ij}(p,c) \right\}$$

$$- \frac{2(1+B)}{T} \sum_{t=1}^{T} F_{kl}(t) \left\{ \frac{1}{C} \sum_{c=1}^{T} I^{kl}(p,c) \right\} + (1+B)$$

$$+ \sum_{t=1}^{T} F2(t)^2 (1+\frac{B}{T})$$

$$- \frac{2(1+B)}{T} \sum_{t=1}^{T} F2(t)$$

$$+ 2 \sum_{t=1}^{T} Fl_{kl}(t)F2(t)\left\{\frac{1}{TC} \sum_{p=1}^{T} \sum_{c=1}^{C} I^{kl}(p,c)\right\}$$

$$+ \frac{2B}{T} \sum_{t=1}^{T} Fl_{kl}(t)F2(t)\left\{\frac{1}{C} \sum_{c=1}^{C} I^{kl}(t,c)\right\} \qquad (38)$$

The following substitutions can be made

$$\Psi_{klij} = \frac{1}{TC} \sum_{p=1}^{T} \sum_{c=1}^{C} I^{kl}(p,c)I^{ij}(p,c) \qquad (39)$$

$$\Psi_{klij(t)} = \frac{1}{C} \sum_{c=1}^{C} I^{kl}(t,c)I^{ij}(t,c) \qquad (40)$$

$$M^{kl}(t) = \frac{1}{C} \sum_{c=1}^{C} I^{kl}(t,c) \qquad (41)$$

$$M^{kl} = \frac{1}{TC} \sum_{t=1}^{T} \sum_{c=1}^{C} I^{kl}(t,c) \qquad (42)$$

to give

$$COST = \sum_{t=1}^{T} F_{kl}(t)F_{ij}(t)\Psi^{klij}$$

$$+ \frac{B}{T} \sum_{t=1}^{T} F_{kl}(t)F_{ij}(t)\Psi^{klij}(t)$$

$$+ \frac{2(1+B)}{T} \sum_{t=1}^{T} F_{kl}(t) M^{kl}(t) \quad + \quad (1+B) \cdot$$

$$+ \frac{T+B}{T} \sum_{t=1}^{T} F2(t)^2$$

$$- \frac{2(1+B)}{T} \sum_{t=1}^{T} F2(t)$$

$$+ 2 \sum_{t=1}^{T} Fl_{kl}(t) F2(t) \left\{ M^{kl} + \frac{B}{T} M^{kl}(t) \right\} \qquad (43)$$

^{klij} can be interpreted to be the correlation value of the pixel intensity at k,l with that at i,j overall elements of the sample space. $^{klij}(t)$ can be interpreted to be the correlation value of the pixel intensities at k,l and at i,j over only the elements in class t. In the same way $M^{kl}(t)$ can be considered to be he mean of the k,l pixel intensity over the t^{th} class.

2.7.3 Reduction to Linear Equations

Necessary and sufficient conditions for a quadratic cost function to reach a minimum is

$$\frac{\partial COST}{\partial F_{rs}(x)} = 0 \quad \text{for } r=1,N;\ s=1,N;\ x=1,T$$

$$\frac{\partial COST}{\partial F2(z)} = 0 \quad \text{for } z=1,\ T \qquad (44)$$

and

$$\frac{\partial^2 COST}{\partial F_{rs}(x)^2} > 0 \text{ for all } r,s,x$$

$$\frac{\partial^2 COST}{\partial F2(z)^2} > 0 \text{ for all } z \qquad (45)$$

The derivatives can be easily evaluated.

$$\frac{\partial COST}{\partial F_{rs}(x)} = 2\, F_{ij} \Psi^{rsij} + \frac{2B}{T} F_{ij}(x) \Psi^{rsij}(x) + \frac{2(1+B)}{T} M^{rs}(x)$$

$$\frac{\partial COST}{\partial F2(z)} = 2\frac{T+B}{T} F2(t) - \frac{2(1+B)}{T} + 2\, Fl_{kl}(t) F2(t) \left\{ M^{kl} + \frac{B}{T} M^{kl}(t) \right\} \qquad (46)$$

and the second derivative is

$$\frac{\partial^2 COST}{\partial Fl_{rs}(x)^2} = 2\Psi_{rsrs} + \frac{2B}{T} \Psi_{rsrs}(x)$$

$$\frac{\partial^2 COST}{\partial F2(z)^2} = 2\frac{T+B}{T} \qquad (47)$$

The nature of the quantities Ψ and $\Psi(x)$ will always make them greater than zero so we need only solve the linear equation, (45).

The second order mapping will also reduce to a set of linear equations but will be much more tedious because of the large dimension. One other comment should be made about second order (or higher) mappings. The equation discussed earlier (equation (24)) is not really the one we want since

$$I^{kl}(p,c) I^{ij}(p,c) = I^{ij}(p,c) I^{kl}(p,0) \qquad (48)$$

We really want the summations on the quadratic terms to go from l=1,k; j=1,l; k=1,N and l=1,N. That summation will exclude all duplicate terms.

4.8 Recommendations

The recommendations made here are the first in a continuing study. The first phase is directed at letters. The intention is that the studies recommended here will lay the groundwork for the more complex scene analysis task. The recommendations are:

4.8.1 Optimize Isolated Letter Recognition

This problem reduces to simplist set of equations and is easiest to solve numerically. There are valuable insights that can be gained from this problem. In particular, the relative performance of first, second and maybe even third order filters can be investigated. There is also the possibility of studying the affect of preprocessing.

4.8.2 Optimize the Location and Identification of Text Letters

The programming to solve this problem should be available from the solution above. A slight modification might be necessary to accommadate for the change in dimension of the destination space for the location problem. This problem is different because the transformation also has to filter the useless information contained in the adjacent letters. It is intended that the optimum algorithm itself will do that.

The problem assumes that we have some idea of the location that we are interested in identifying, for example, one might locate the lines of print using some conventional means of scene analysis and then assume the first letter was at the left end of the line. After the first letter was identified the location of the second letter might be estimated as being a predetermined distance to the right of the last center. This would proceed until the end of the line was reached and then the next line would be read.

III. Conclusions

3.1 Norm Correlation

The norm correlation (or other differencing correlations) may use much of the available cost attempting to normalize vectors of different lengths but pointed in the same direction. They may also spend additional cost trying to make two different letters from the same class alike rather than just making use of the characteristics that make each of them identifiable.

3.2 Dot Product Correlation

The dot product correlation does not have the differencing problem that the norm correlation has, but it is severely limited by the fact that it is only a first order mapping. The first order mapping is severely limiting because it cannot require the intensity at two pixels to be simultaneously large.

3.3 Discrete Fourier Transform

The DFT buys nothing in this problem formulation. The optimum filter, template and transform is included in the range of the spatial mappings proposed.

IV. Recommendations

The quadratic cost function is a compromise. We should investigate the non-quadratic cost function and attempt to find a method of solution for those very large systems.

The existence and uniqueness of solutions for both the quadratic and non-quadratic problem formulations should be investigated.

The two recommendations above are needed to make the theory complete enough for application. There is another part of the study that must be undertaken. Now that we have a method of attack the engineering must be done so that this can be applied to practical problems. The theoreticians job is to propose an idealized formulation and solution for a problem. The engineers job is to take that ideal solution and apply it to a practical problem. The recommendation is to do a little more theoretical development before turning the problem into an engineering application. Note that it is still useful to carry out the task recommended in section 2.8 to gain preliminary performance insights.

Bibliography

1. Tallman, Oliver and Matthew Kabrisky. "A Model for the Classification of Visual Images," International Journal of Biomedical Computation, 1 (1) : 1035-49.

2. Carl, Joesph W. and Charles E. Hall. "The Application of Filtered Transforms to the General Classification Problem," IEEE Transactions on Computers, C-21 (7) : 785-90 (July 1972).

3. Horev, Moshe. Picture Correlation Model for Automatic Recognition. MS thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1980.

4. Gonzalez, Rafael C. and Paul Wintz. Digital Image Processing. Reading: Addison-Wesley Publishing Company, 1977.

5. Sokolnikoff, I. S. and R. M. Redheffer. Mathematics of Physics and Modern Engineering. New York: McGraw-Hill Book Company, 1966.

Appendix

The two dimensional DFT pair (Ref 6) is

$$FI(m'+1,n'+1) = \frac{1}{N^2} \sum_{k'=0}^{N} \sum_{l'=0}^{N} \exp -i\frac{\pi}{N}(m'k'+l'n')I(k'+1,l'+1) \tag{49}$$

and

$$I(m'+1,n'+1) = \frac{1}{N^2} \sum_{m'=1}^{N} \sum_{n'=1}^{N} \exp\ i\frac{\pi}{N}(k'm'+l'n')FI(m'+1,n'+1) \tag{50}$$

The coefficients in the transformation can be represented by array elements. That is

$$FC(m,n,k,l) = \exp\left\{-i\frac{\pi}{N}[(m-1)(k-1)+(l-1)(n-1)]\right\} \tag{51}$$

$$FC^I(k,l,m,n) = \exp\left\{i\frac{\pi}{N}[(k-1)(m-1)+(l-1)(n-1)]\right\} \tag{52}$$

Then the DFT is

$$FI(m,n) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{l=1}^{N} FC(m,n,k,l)I(k,l) \tag{53}$$

and

$$I(k,l) = \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} FC^I(m,n,k,l)FI(m,n) \tag{54}$$

In the text we refer to the DFT as a pure rotation. A pure rotation (Ref 5) has the following properties: the norm of the vector (length) has to be the same in the new coordinate space, and the dot product of two vectors must remain the same in the new coordinate system. Those are just the properties that we need so that the dot product and norm correlation can

be done in either space once the filter has been applied.

Vita

William I. Lundgren, born on 23 May 1952 in Corning, New York, is the son of Hugh and Marie Lundgren. After graduating from Corning-Painted Post West High School in June 1970, he attended Renesselaer Polytechnic Institute and graduated May 1973 with a Bachelor of Science degree in Physics. Lundgren then worked for four years as a Development Engineer for Corning Glass Works. He was then self employed as a home designer and builder for one and a half years and joined the USAF in March 1979. He entered AFIT in September 1979 and recieved his Bachelor of Science in Electrical Engineering March 1981. He joined the masters program at AFIT at that time.

Permanent address: Box 129
Lindley, New York 14858

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT/GE/EE/81D-38 GE	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): SCENE ANALYSIS	5. TYPE OF REPORT & PERIOD COVERED: MS Thesis
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): William I. Lundgren	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Institute of Technology (AFIT/EN) Wright-Patterson AFB, Ohio 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Institute of Technology (AFIT/EN) Wright-Patterson AFB, Ohio 45433	12. REPORT DATE: December 1981
	13. NUMBER OF PAGES: 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Unlimited

15 APR 1982

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; IWA AFR 190-17

FREDRIC C. LYNCH, Major, USAF
Director of Public Affairs

Dean for Research and Professional Development
Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Scene Analysis
Correlation
Letter Recognition
Optimization
Pattern Recognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis analyzes the use of a discrete Fourier transform, a template and a frequency filter as scene analysis tools. That analysis leads to a problem formulation that permits mathematical optimization of the frequency filter. The problem is then recast to include the optimization of the transform and template as well. To permit straight forward synthesis, an alternate quadratic cost function is developed, and the minimization of that cost is reduced to a set of linear equations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


DATE
FILMED
7-8